DOCUMENT RESUME

ED 041 949 TM 000 044

AUTHOR Jaeger, Richard M.
TITLE Evaluation of National Educational Programs: The Goals and the Instruments.
INSTITUTION Bureau of Elementary and Secondary Education (DHEW/OE), Washington, D.C.
PUB DATE Mar 70
NOTE 20p.; Paper presented at the annual meeting of the American Educational Research Association, Minneapolis, Minn., March 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$1.10
DESCRIPTORS Achievement Tests, Data Collection, Evaluation, *Evaluation Criteria, Evaluation Techniques, Federal Aid, *Federal Programs, Item Sampling, Measurement, *Measurement Instruments, Measurement Techniques, *National Programs, National Surveys, *Program Evaluation, Research Design, State Federal Aid, State Federal Support, Statistical Surveys, Testing
IDENTIFIERS Joint Comprehensive Evaluation System, *U. S. Office of Education

ABSTRACT

A new Joint Comprehensive Evaluation System for the assessment of 15 different federal programs has been developed by the U.S. Office of Education. In this system these diverse program services will be thought of as resources available to meet the needs of critical target groups. Using this approach, a set of nine crucial questions that need to be answered in program management have been developed. The evaluative design for finding the answers to these questions proposes to use the individual pupil as the unit of analysis. By use of sample survey methods and multiple matrix sampling where different individuals complete different samples of test items it will be possible to collect comparable and generalizable data without putting an undue testing burden on any one student. The data collection instruments are discussed in some detail. (DG)

ED041949

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

EVALUATION OF NATIONAL EDUCATIONAL PROGRAMS: THE GOALS AND THE INSTRUMENTS

by

Richard M. Jaeger
Chief of Evaluation Design
Bureau of Elementary and Secondary Education
U.S. Office of Education

Presented at the
1970 Annual Meeting:
American Educational Research Association
Minneapolis, Minnesota

430 000 ERIC

The Problem

Many researchers have attempted to define evaluation; some by stating what it is not, others by stating what it is. To Stake and Denny (1968), "Evaluation is not a search for cause and effect, an inventory of present status, or a prediction of future success. It is something of all of these but only as they contribute to understanding substance, function and worth." Perhaps the evaluation program which we are describing today can best be judged by the criterion proposed by Hemphill (1968): "... the worth of an evaluation study is to be found in its contribution to a rational decision process..."

Clearly, what we are about in the development of a Joint Comprehensive Evaluation System, is an attempt to make rational a complex set of decision processes.

The process by which Congressional intent is transformed into educational practice has been described by others participating in this symposium. If we are to produce a successful end, we must demand management, skill and informed decision-makers at every stage of the transformation process -- from the program administrators in the Federal office, to the grants managers in State Departments of Education, to project developers in local school systems, to the principals and teachers who, in the final analysis, make education work. The evaluation system which we are describing today seeks to provide needed information for all of these decision-makers, although the components I shall discuss will primarily

DOCUMENT FILMED FROM BEST AVAILABLE COPY

serve those in State Departments of Education and the Federal office.

To build an evaluation system which will contribute to a rational Federal and State decision process, we began by analyzing the decisions to be made. For each of the legislative programs administered by the Bureaus of Elementary and Secondary Education and Vocational and Technical Education, legislation, guidelines, regulations and administrative criteria were carefully analyzed to better understand the process through which Federal educational funds were transformed to local educational programs. Having built a decision model for each program, there began a most important phase of the developmental work; perhaps the activity which allows this evaluation program to be termed "Joint State/Federal". Through an iterative process of suggestion and modification, each of the programs' administrative officers in the cooperating States and the Office of Education worked to define the information base necessary to rational and effective program management. Additionally, there evolved a data base to be used in the critical task of informing the publics of Federally supported education programs of the status and progress being realized.

To state the major points of decision and questions of policy associated with each of the 15 legislative programs we seek to evaluate would require at least the balance of this symposium. Fortunately, the

relative similarity of program administrative processes permits the description of a common set of information required for effective State and Federal program management:

Each of the Federally supported educational programs seeks to meet a set of needs, defined either by specific activities for which funds may be expended, or by designation of a group of pupils and education professionals for whom services are intended. In either case, it is managerially sound to consider program services as resources available to meet the needs of critical target groups. The first question of importance to program management can thus be stated as follows:

1. What is the size of critical target groups of pupils and education professionals and what is their demography?

Answers to this question provide measures of the global need for the services which legislative programs authorize, and State and National pictures of the demographic concentration of those in need of service.

Answers to two additional questions are necessary to guage the adequacy of present educational programs:

2. What is the size of critical target groups being served under current legislative programs?

3. What is the size of critical target groups not being served under current legislative programs?

To derive indices of the efficiency of programs in reaching critical target groups, State and Federal program managers must know:

4. What proportion of those needing services provided through present educational programs are receiving such services?

5. What proportion of those receiving services under present educational programs are not in need of such services?

Many of the legislative programs we seek to evaluate presume from the outset that success will require novel and innovative approaches to solving the problems they seek to resolve. Thus Title I of ESEA, in its Declaration of Policy of the Congress, notes the "special educational needs of educationally deprived children" and provides funds to local education agencies to "expand and improve their educational programs". Similarly, the Declaration of Policy of the National Defense Education Act speaks of "additional and more adequate educational opportunities". Title III of that Act, for strengthening instruction in critical subject areas, clearly requires actions beyond "more of the same thing". It is of critical importance therefore, that State and Federal program managers know the character of educational services currently being provided under Federally supported programs. The sixth question to be answered is therefore:

6. What is the nature and content of services being provided through Federally supported educational programs, and how do these services compare to those being provided under regular programs supported through State and local funds?

To determine the efficiency of educational programs in directing needed services to critical target groups, it is necessary but insufficient to note the number or proportion of target group members being served. The breadth of services authorized under programs such as Title I, and III of ESEA precludes the determination of program efficiency on the basis of global proportions. State and Federal program managers must know the extent to which specifically needed services are being adequately provided and efficiently directed. We cannot count as success the provision of health services to a healthy child in a Title I school who is functionally illiterate. The seventh question to be answered is then:

7. How well are critically needed educational services being directed to those most in need of such services?

To inform the Congress of needed modificatons in Federal educational policy and to modify guidelines and regulations for more effective program operation, program managers at State and Federal levels require assessments of the overall success of Federally supported educational programs in meeting their specified objectives. Some programs, such as Title II of ESEA, can be termed successful if authorized services are rationally destributed in relation to need and in accordance with legislative criteria. Other programs, such as Title I of ESEA, require

demonstration of progress in solving national educational problems of major scope before success can be claimed. In either case, we must seek answers to the question:

8. How effective are Federally supported educational programs in meeting their stated objectives?

Most of the programs we seek to evaluate place great planning and management responsibility upon State Departments of Education. Under three titles of ESEA and two titles of NDEA, State Departments of Education act as grants managers in approving proposals submitted by local education agencies. Four of these programs require States to prepare comprehensive plans for the disposition of funds in accordance with the findings of statewide needs assessments. To function effectively in the awarding of grants State program officers must be able to identify those proposed projects which have the greatest probability of success. Frequently, educational research findings provide theoretical bases for setting success expectations, but do not afford the assurance of project demonstration under field conditions. State managers require documentation of successful and unsuccessful projects, activities, and treatments, to build a reference library for grants award decisions. The ninth question to be answered through the Comprehensive Evaluation System is thus:

9. What projects, activities and treatments show, through field demonstration, high probabilities of success in meetings stated educational objectives?

These then are questions to be answered by the Joint Comprehensive Evaluation System. In a homely manner they may be summarized as follows:

Who is to be served?

Who is (and is not) being served?

How efficiently are services being provided?

What kinds of services are being provided?

How well are services being directed?

How effective are Federally supported educational programs?

What techniques of educational intervention work?

Finding the Answers - The Design

To answer the evaluative question we have identified requires a complex system of data collection, analysis and reporting. Moreover, the system must be based in a unified design of research.

In designing the 1968 evaluation of Title I, ESEA, the Office of Education employed a research approach unprecedented in national evaluation studies. Unlike previous Title I evaluations which focused on schools, school districts or educational projects, the 1968 Survey on Compensatory Education used individual pupils as units of analysis. The result was a better understanding than ever before, of the composition of the Title I pupil population, the nature of the services being provided under Title I, and the efficiency of the Title I program in directing compensatory services to educationally and economically deprived pupils. These findings alone provided answers to six of the ninequestions we have listed as objectives of State and national program evaluation.

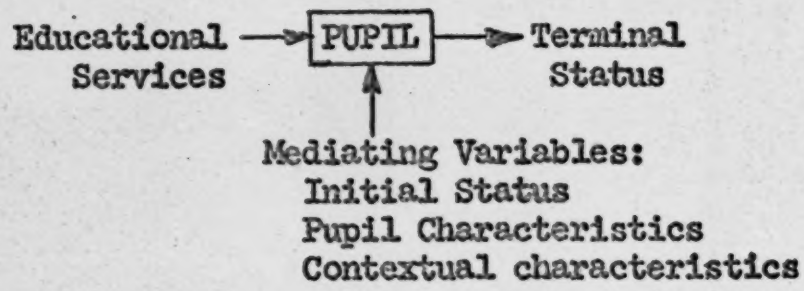
On the basis of previous success, a model which uses pupils as units of analysis will be central to the design of the Joint Comprehensive Evaluation System. Additionally, some components of the system will use projects or activities as units of analysis, in order to improve the efficiency of previous evaluations and to increase the depth and reliability with which educational services can be defined.

In employing a pupil-centered evaluation model, we shall secure four classes of data to provide bases for description and relational analysis. To answer the question "Who is to be served?", we shall secure data on the social status, economic status and educational status of individual pupils. To answer in part the question "What kinds of services are being provided?" we shall secure data on the participation of individual pupils in an array of Federally supported educational projects and activities. By relating these two classes of data, we shall derive answers to the questions "Who is being served?", "Who is not being served?", "How efficiently are services being provided?" and "How well are services being directed?". Not all of our questions on efficient direction of services can be answered through data on individual pupils. We must also determine the educational contexts to which services are directed. In evaluating Title I, for example, it is important to determine the extent to which compensatory programs are directed to schools with high concentrations of economically disadvantaged pupils, as well as determining the characteristics of individual program participants. We shall therefore secure data on the character of institutions -- a set of contextual variables describing the social, economic, educational and ethnic compositions of schools and school systems.

To answer questions on the effectiveness of some programs and to build a catalog of successful educational projects, we shall secure data on the academic status of pupils both before and after their exposure to federally

supported educational projects. A unique component of our evaluation program is a set of pupil status measures through which we shall obtain generalizable group achievement information. The efficient techniques of multiple matrix sampling to be employed in collecting these data will be described in greater detail. The questions "How effective are Federally supported educational programs?" and "What techniques of educational intervention work?" will be answered by relating data on pupil status, pupil and contextual characteristics, and educational services. Multivariate correlational techniques will be employed, with terminal pupil/status as dependent variables, characteristics of educational services as independent variables and pupil and contextual characteristics as mediating variables.

Schematically, the pupil-centered model may be portrayed as follows:



This then is our research model, I shall next describe the instrumentation procedures through which the model will be employed.

Finding the Answers - The Instrumentation

The Comprehensive Evaluation System will apply the pupil-centered evaluation model through a series of sample surveys employing four types of data collection instruments. Surveys will be designed to yield high precision estimates of variables basic to the pupil-centered model. Data obtained will generalize to national populations of pupils, teachers, schools and school districts.

We have termed one set of questionnaires "Pupil Centered Instruments". The Pupil Centered Instruments are used to build four relatable files of information on school districts, schools, teachers and pupils. The School District Questionnaire will build upon the basic program-accounting information secured through the Comprehensive Program Information Report already described. It will obtain a more detailed picture of the educational program functions conducted centrally within school systems. Title I, ESEA, among other programs, contains provisions for the training of professional personnel and the involvement of parents and community members in the planning and implementation of programs. Both of these activities are generally administered by central school system offices. The School District Questionnaire will secure information on the participants, substance and activities of training and community involvement programs for purposes of program description and, in conjunction with other information, explore the effectiveness of such programs. The school district instrument will also provide limited information on the district administration of Federally supported programs. The

School Questionnaire will be completed under the direction of principals, and will provide vital information on services and instructional resources available to students in addition to data on the social, ethnic, economic and academic composition of student bodies. These data will provide a basis for deriving estimates of need for educational services and resources in schools across the nation. By relating information on the availability of Federally supported educational programs to data on student body characteristics and other resources available in schools, a critical link in the direction of needed services to individual pupils will be examined. Data obtained from the School Questionnaire will also be used as mediating variables in multivariate analyses of the relationship between pupil participation in Federally supported programs and changes in pupil behaviors. The Teacher Questionnaire will provide five classes of information. First, information on the background, qualifications and training of teachers will be obtained. These data will be used to assess needs for additional training, to examine the efficiency of direction of existing training programs, and, when related to other variables, to search for evidence on the effectiveness of training programs. The second class of data provided by teachers will concern the organization and composition of classes. These data will be used to determine the extensiveness and effects of pupil grouping by academic status, social status and ethnicity. Additionally, these data will provide indication of the efficiency of direction of

Federally supported educational services to classroom groups. The third and fourth classes of data will concern methods of teaching and programs of instruction. In addition to providing a basis for the examation of innovation, novelty and improvement in Federally supported programs, data on regular programs of instruction are necessary to the analysis of Federal program effectiveness. Differences in regular programs of instruction across schools and classes must be examined and statistically so as not to confound analyses of Federal program effectiveness. The fifth class of data to be provided by teachers includes teacher perceptions of classroom climate, adequacy of instructional resources and appropriateness of instructional resources. Additionally, data on teacher attitudes will be secured. These data will permit further examination of the quality of Federally supported programs and will provide important information on an immediate effect of professional training programs. Teacher attitude information will also be used as mediating variables in examinations of program effectiveness. The Pupil Questionnaire will provide data on individual children indispensable to the pupil-centered evaluation model. Five classes of data will be secured. The first set of data will allow classification of pupils as to age, sex, transiency, attendance and special educational categories. The second class of pupil data will allow the development of indices of social, economic, ethnic and academic status for individual pupils. These indices are vital to analyses of national needs

for educational services and the efficiency of Federally supported programs in providing services in accordance with pupil needs. The third and fourth classes of data will indicate the extensiveness and intensity of pupil participation in Federally supported academic and ancillary programs. These data permit one to follow Federally supported educational programs to their ultimate targets. The importance of this component of the Comprehensive Evaluation System cannot be overstated. Hollingshead, Warner, Sexton and other educational sociologists, over a period of four decades, have demonstrated the dangers of equating availability of educational resources with provision of educational services. A number of the critical targets of Federal programs -- the economically disadvantaged, the educationally disadvantaged, the children of agricultural migrant families -- have been shown by these sociologists to be the least likely consumers of specialized educational resources, in the absence of explicit program participation criteria. The fifth class of pupil data includes teacher reports on important criterion behaviors for individual pupils. In the wake of growing professional recognition of the necessity but insufficiency of standardized achievement tests as indicants of educational success, these variables will provide a broad base for examination of program effectiveness in the socialization, motivation and self actualization of pupils.

While the Pupil-Centered Instruments form the heart of the Comprehensive Evaluation System, they do not provide vital elements of data secured

more efficiently through other sources. To move beyond an analysis of the global effectiveness of Federally supported programs to analyses of the effectiveness of locally implemented projects requires considerable information on the objectives, resources and operations of those projects. Additionally, such information is required to answer questions on the improvement of instructional services which results from Federal educational support, and the appropriateness of educational services to the needs of participants. A survey instrument which utilizes Federally supported projects as units of analysis will be used to secure these critical data. This instrument has been termed a "Project Descriptor", and will be completed by knowledgeable project administrators in school district offices and schools. By supplying information on the participants, objectives, reources, processes and organization of Federally supported projects, the Project Descriptor will provide a basis for building a reference library for State grants managers. When data from the Project Descriptor are integrated with information from the Pupil Centered Instruments and criterion measures yet to be described, a resource information bank on effective educational projects can be assembled.

The use of sample survey methods in national evaluation studies requires meticulous attention to the structure of samples. Since effective evaluation requires collection of a wealth of information, the efficiency of sample design is critical to study feasibility. Utilizing a complex multistage sampling design, the Comprehensive Evaluation System will

provide nationally generalizable information by securing data in 830 of the nation's 19,000 school districts. To further improve sampling efficiency, a complete reference file of schools with Federally supported educational projects is being developed in cooperation with State Departments of Education and the Office of Education's National Center for Educational Statistics. This Project Reference File will provide a minimum of information on the existence of projects by source of support and target grade in order to build a sampling frame for maximal efficiency in the selection of schools. The data derived from the Project Reference File will also permit unprecidented analyses of the extent to which the various Federal educational programs are used by local school district managers to provide services to a common group of pupils.

Since the inception of program evaluations by the Office of Education, securing comparable and generalizable data on pupil achievement has been the bane of effectiveness analyses. In the first years of Title I evaluation, the diversity of evaluation bases employed by States and school districts thwarted attempts to examine national program effectiveness. The 1968 and 1969 Surveys on Compensatory Education, utilizing consistent instrumentation in a national sample of schools and school districts, provided the first hard data on pupil needs, educational services and program efficiency. Unfortunately these Surveys could not use common tests of pupil achievement, and data collected were those available in schools. The loss of common

achievement data again precluded assessment of a critical dimension of Title I program effectiveness. Only nine percent of the pupil questionnaires secured in the 1968 Survey contained analyzable achievement change data. Unfortunately, the size and distribution of this pupil sampling did not allow national generalization of findings on pupil achievement. To overcome this persisting problem, the Comprehensive Evaluation System will employ a method of testing at the forefront of psychometric theory. Multiple matrix sampling, a procedure by which different individuals complete different samples of test items, is based on an analytic development by Lord (1955). Cronbach suggested the use of matrix sampling in the evaluation of instructional programs in 1963. Since then, matrix sampling has been employed experimentally in the development of test norms and has been used most extensively in the National Assessment Program. The procedure is ideally suited to large scale evaluation programs. In the Comprehensive Evaluation Program, pupils in the classes to be surveyed will complete a series of sampled tests, each requiring no more than ten minutes of pupil time. The resulting data will provide reliable achievement statistics for groups of pupils, both participants and non-participants in Federally supported programs, but will not provide reliable data for individuals. Our evaluative use of achievement data requires inferences on the performances of groups rather than those of individuals. Hence the lack of reliable data for individuals is unimportant. The use of matrix

sampling allows minimal disruption of classes and minimal investment of testing time to secure consistent achievement data. Pupil Status Measures in the construct areas Basic Verbal Status and Occupational Cognizance for pupils in grades four and eleven have already been developed and tested on 300 children. The results are very encouraging. With test means for groups in the range 34 to 49, standard errors of means ranged from 24 hundredths of an item to 56 hundredths of an item. Thus coefficients of generalizability, were in the range .83 to .93. For individuals, test-retest reliabilities ranged from .61 to .82. The pretest of these instruments also showed discrimination with respect to the socio-economic composition of schools usually associated with standardized achievement tests. However, a preliminary testing in schools with 90 percent poor Chicano children produced no indications of ethnic or language bias. Data resulting from application of these common status measures, when combined with information from the Pupil-Centered Instruments and the Project Descriptor, will allow determination of the efficiency of direction of Federally supported educational services to academically needy pupils. More important, these data will provide the first comprehensive basis for investigation of a critical criterion of program effectiveness.

Before concluding, I should like to tell you of another project which may provide a method of securing generalizable achievement test data. The patterns of test utilization in U.S. elementary schools determined

from the 1968 Survey on Compensatory Education indicated extensive use of six achievement test batteries. At present, results obtained from administration of these tests cannot be combined. The tests differ somewhat in content. Moreover, they were standardized on decidedly different norm groups. Last year, the Office of Education contracted for a study of the feasibility of restandardizing the reading subtests of these batteries, at the upper primary grade levels. If these tests could be restandardized on a common nationally representative sample of pupils, one of the major obstacles to combining test results would be removed. Additionally, if one of these tests could be used as a reference "anchor", scores on different tests might be equated with reliability sufficient for evaluative applications. The results of the feasibility study were quite encouraging. A full scale anchor test study is now under consideration.

These then are the goals and the instruments. With a monumental investment of energy on the part of educators in the States, local school systems and the Federal Office, a comprehensive system for the evaluation of Federally supported education programs has been conceived. With the contientious aid of consultant scholars and many in private industry, the further efforts of these individuals will see to fruition a system which meets Dr. Hemphill's criterion. We shall indeed make rational a most complex decision process.